

**“The power of the voice”: virtual voice assistants, voice commands
and contracting by means of voice**

**“O poder da voz”: assistentes de voz virtuais, comandos de voz e
contratação por meio da voz**

Julia Ammerman Yebra

Postdoctoral researcher “Margarita Salas”

Facultad de Derecho/Law Faculty, Universidade de /University of Santiago de
Compostela

Avda. Dr. Ángel Echeverri, s/n, 15782, Santiago, Spain

julia.ammerman@usc.es

<https://orcid.org/0000-0001-7858-0541>

September 2022

ABSTRACT: Voice is a feature that individualizes and identifies us. Also, a voice is a biometric data, which means that once a voice has been processed, its owner can be identified. This allows us to state that the voice can be used as a means to authenticate the signature of an electronic contract. Therefore, biometric voice recognition techniques will make it possible to identify the owner of the voice, who will thus be giving his contractual consent. The issue could become more complicated when not only the “voiceprint” is used as a signature, but also the counterpart is a virtual voice assistant. These assistants, designed to both “answer” and “listen”, should contain strong privacy policies regarding the processing of the voices interacting with them. The purpose of this paper is to analyze the consent given both for electronic contracting by voice and for the processing of these voices by virtual assistants.

KEY WORDS: voice; electronic contract; biometric recognition; voice virtual assistant; contractual consent; privacy policies.

RESUMO: A voz é uma característica que nos individualiza e nos identifica. Além disso, a voz é um dado biométrico, o que implica que, uma vez processada uma voz, o seu proprietário pode ser identificado. Isto permite-nos afirmar que a voz pode ser utilizada como meio para autenticar a assinatura de um contrato electrónico. Assim, as técnicas biométricas de reconhecimento de voz permitirão identificar o proprietário da voz, que dará assim o seu consentimento contratual. A questão pode tornar-se mais complicada quando não só a “impressão vocal” é utilizada como assinatura, mas também a contraparte é um assistente de voz virtual. Estes assistentes, concebidos tanto para “responder” como para “ouvir”, devem conter fortes políticas de privacidade relativamente ao processamento das vozes que interagem com eles. O objectivo deste trabalho será analisar o consentimento dado tanto à contratação de vozes electrónicas como ao processamento destas vozes por assistentes virtuais.

PALAVRAS-CHAVE: voz; contrato electrónico; reconhecimento biométrico; assistente de voz virtual; consentimento contratual; políticas de privacidade.

TABLE OF CONTENTS:

1. Introduction: the intangibility of the voice vs. the growing power of the “spoken word”

1.1. What can neither be seen nor touched

1.2. The power of the “spoken word”

2. Voices and contracts: the voice in the different phases of contracting

2.1. The voice in the initial phases of the contract

2.2. The voice in the phase of conclusion of the contract

a) Verbal contracts

b) The signing of contracts by means of the voice

3. User interaction by means of voice and virtual voice assistants

3.1. Voice processing as biometric data

3.2. Analysis of the consent given by the users to the processing of their voice data

a) When the processing of voice data is necessary for the performance of the contract. The case of virtual voice assistants

b) When the processing of voice data is not necessary for the performance of the contract. The case of “voice data” clauses in other services

4. Final conclusion

Bibliography

1. Introduction: the intangibility of the voice vs. the growing power of the “spoken word”

1.1. What can neither be seen nor touched

The voice cannot be seen or touched. It is intangible. But it is heard, perceived by the ears. The voice can be sung, recited, spoken. The word can be shouted or whispered. Spoken in urgency or savoring each syllable. And remain the same, unique voice of an individual. It may be more or less known to the public, but it will always retain its uniqueness, for it is a personality trait that identifies each person. In the words of Danièle Huet-Weiller, the voice is “a second face, a substitute for physical presence that makes it possible to identify the individual without the aid of the eye”¹.

However, this identifying function of the voice, of individualizing people, has not yet penetrated very deeply into society. The intangibility of the voice is perhaps the reason why it has never received the same attention as, for example, the image of a person. “A picture is worth a thousand words”. Indeed, the visual impact is usually greater than the sound, or at least our society has come to understand it that way. In the same vein but in another context, recently an organist said that Spain was a country of light, “with a visual culture, which enters through the eyes and not through the ears”. And perhaps that is why organs, instruments that produce sound, heritage that makes music, were often forgotten by institutions when it came to restoring them². Transferred to the field of the voice, it is significant that still in countries like Spain there is no consensus among the legal doctrine on whether or not the voice constitutes an autonomous personality right, as is the case with the image or the name³. This lack of attention to the voice contrasts with the growing value that this personality feature is acquiring in different digital interactions, as we shall see below.

Although there is no consensus on its establishment or not as a personality right, what has been pointed out from a legal point of view is that the second function of the voice —the first would be the communicative or expressive— is that of a way of representation and identification of persons⁴. This is a consequence of the fact that the voice, as sound material,

¹ DANIELLE HUET-WEILLER, “La protection juridique de la voix humaine”, *Revue Trimestrielle de Droit civil*, 1982, p. 497.

² Interview with the organist Roberto Fresco in the newspaper La Región, June 10, 2022. <https://www.laregion.es/articulo/celanova/somos-pais-luz-nivel-patrimonial-organo-hace-musica-entra-ojos/202206092127341137759.html> (29.08.2022).

³ In other legal systems, the following authors have defended a right to one’s own voice as an autonomous personality right: in Italy, ADRIANO DE CUPIS, *I diritti della personalità*, Vol. IV, T. I, Milán, Giuffrè, 1959, p. 295 and GIORGIO RESTA, *Autonomia privata e diritti della personalità*, Nápoles, Jovene Editore, 2005, pp. 167-169; in France DANIELLE HUET-WEILLER, “La protection juridique de la voix humaine”, *Revue Trimestrielle de Droit civil*, 1982, pp. 497-513; in Germany ANKE SCHIERHOLZ, *Der Schutz der menschlichen Stimme gegen Übernahme und Nachahmung*, Baden-Baden, Nomos, 1998; in Portugal RAQUEL GUIMARÃES, “A tutela da pessoa e da sua personalidade: algumas questões relativas aos direitos à imagem, à reserva da vida privada e à reserva da pessoa íntima ou direito ao carácter”, *A tutela geral e especial da personalidade humana* [ebook], Centro de Estudos Judiciários, 2017. In Spain we have recently defended the protection of the voice as an autonomous personality right in JULIA AMMERMAN YEBRA, *El derecho a la propia voz como derecho de la personalidad*, A Coruña, Colex, 2021.

⁴ PABLO FERRÁNDIZ AVENDAÑO, “Protección del artista frente a la imitación de su voz”, *Actualidad Civil Jurisprudencia*, no. 21, Wolters Kluwer, 2010, p. 3.

is broken down into three acoustic parameters —timbre, frequency and intensity— which together result in each person having a voice different from the rest. Timbre is perhaps the most descriptive parameter of the voice, since it allows distinguishing two sounds of equal frequency, duration and intensity⁵. It provides the listener with valuable information about the owner of the voice, allowing the generation of an approximate profile of the person based on age, gender and even physical appearance. Vocal frequency refers to the intensity peaks —or “formants”— in which most of the sound energy of a given sound is concentrated, thus allowing to distinguish the sounds of human speech, especially vowels. The frequency of vibration (so many vibrations per second) determines the height: the slower the vibrations follow one another, the deeper the sound will be, and the faster the vibrations, the higher it will be⁶. Finally, the intensity depends on the circumstances in which the voice is used, and can range from an almost inaudible mumbled voice to an intensity with values that are difficult to tolerate with the human ear (in the order of 110 decibels)⁷.

In our opinion, this identifying function of the voice is not limited to individualizing each person as do fingerprints or the iris of the eye. Although we already speak of the “voice print”, the identifying function of the voice goes beyond this, *linking itself* to dignity and the free development of the personality. In this sense, we understand that we are dealing with a personality feature that defines us as persons and makes us recognizable as individuals, bearing values and culture, in a similar way as image and name do. Although there are voice parameters determined by our genetics, there are others that we modulate according to our personality and way of being, partly due to the expressive function of the voice. The fact that artists must expressly authorize the dubbing of a recording of their own when the same language used by them is to be used, or their own language (or one of their own languages) when the recording was made in another language, is significant of the effect that assigning a voice not corresponding to that of the person represented can have on the personality. Also revealing are the initiatives that promote “donate your voice” for patients who have lost or will foreseeably lose the ability to speak: having the voice of a known person, or even their own voice, in the voice synthesizers used by these people, allows them to preserve part of their personality, avoiding robotized voices and alike for all patients with such diseases⁸.

⁵ Timbre is the consequence of the collision of air with the oral-nasal cavity, the lips, the tongue and the teeth, thus determining the form that ends up adapting a voice. In the words of GUY CORNUT, *La Voix*, Paris, Presses Universitaires de France, 1990, p. 43, “it is the result of the transformation and shaping of the laryngeal sound by the resonance cavities”.

⁶ PERCY A.SCHOLES, / JOHN OWEN WARD / DEVOTO, Oxford Dictionary of Music, T. I and II, Barcelona, Edhasa / Hermes / Sudamericana, 1984, p. 50 (“Acoustic” voice, Oxford Dictionary of Music). In the spoken voice, tone pitch oscillates around an average frequency that is almost always the same for a subject having a normal conversation, but differs between children, higher pitch (usually between B2 and E3), women, a little lower (between G2 and D3) and men, quite a bit lower (between G1 and D2).

⁷ GUY CORNUT, *La Voix*, ... cit., pp. 41-42.

⁸ The processes for the “creation” of these voices consist of the recording of about one hundred phrases (which are not random, as they contain all the sounds necessary to construct these voices), creating some with a more natural timbre. The researchers at the Universidad del País Vasco who have developed this project (“Aholab”) aim to humanize communication between patients who have lost the ability to speak, avoiding “robotic voices without personality”. On their web page you can see all their projects: <https://aholab.ehu.es/aholab/> (29.08.2022).

1.2. The power of the “spoken word”

On the other hand, since the invention of sound recording systems, the voice has become increasingly important. As early as the beginning of the 19th century, tests appeared to record sounds by means of a needle attached to a membrane that vibrated sympathetically with the sounds produced on it, thus allowing the tip of the needle to impress a plate moving at a certain fixed speed. Subsequently, it was possible to replay what was previously recorded, which is known as the “reversibility principle” of the recording process. Thus, both on one side of the Atlantic and on the other, Cros and Edison invented the first phonograph⁹, the latter being credited with the construction of the first apparatus. In the words of Cros “Mr. Edison was [...] the first to reproduce the human voice. He has accomplished something admirable”¹⁰.

At first the invention was considered, even by its creators, as something “little more than a toy”, or at best a device for office use. Perhaps for this reason, musicians were slow to recognize the value of the phonograph, a situation that changed when, at the beginning of the 20th century, singers allowed their voices to be recorded. Thus, the tenor Enrico Caruso made his first recordings in 1902, Charles Santley in 1904, or Nellie Melba in 1905¹¹.

The possibility of voice fixation seems not to have escaped the attention of one of the German jurists credited with formulating the specific category of personality rights. We are talking about Karl von Gareis (1844-1923), who —referring to the category— said that “the series of these alleged rights is not yet closed, for one can speak of a right to one's own voice”¹². It was not until almost a century later that a French jurist, Ivainer, wrote an article in 1966 entitled “*Le magnétophone, source ou preuve de rapports juridiques en droit privé*”, and suggested, probably influenced by the discussions within the commission for the reform of the *Code Civil*, the addition of art. 162 of the preliminary draft of the reform —according to which any person could prohibit the publication, exhibition or use of his image— by adding analogous protection for the voice: that any person could also prohibit “the recording or use of his voice”¹³.

Without going into the evolution that this doctrine meant for the theory of personality rights, what we are now interested in emphasizing is that the “spoken word” can now be analyzed,

⁹ Regarding the term phonograph or gramophone, it seems that the first one was coined by Edison in 1877 to name his device, since he called his records “phonograms” (“written voice”, from the Greek *phoné*, voice, and *graphein*, to write). Since the brilliant inventor claimed for himself exclusive rights to the name, the inventor Berliner, who improved the device, called it “gramophone” (which seems to be an inversion of “phonograph”). PERCY A. SCHOLLES / JOHN OWEN WARD / DEVOTO, *Oxford Dictionary...* cit. p. 600.

¹⁰ PERCY A. SCHOLLES / JOHN OWEN WARD / DEVOTO, *Oxford Dictionary...* cit, p. 595, reproducing a curious excerpt from *The Musical Times* of November 1887, in which, speaking of Edison's new invention, they asked: “Will singers and performers wish to sing and play before the 'receiver,' and send examples of their ability all over the world? Will Rubinstein or little Hofmann make world tours by phonograph, quietly sitting at home while preparing new copies and while their agents are traveling and exhibiting them (...)” and later in 1898: “In answer to your question about the new Edison phonograph, we can tell you from personal experience that it is a really wonderful instrument (...)”.

¹¹ PERCY A. SCHOLLES / JOHN OWEN WARD / DEVOTO, *Diccionario Oxford...* cit., p. 598.

¹² Quoted by FRANCESCO FERRARA in his *Trattato di Diritto civile italiano de 1921*, and cited by PEDRO RUIZ Y TOMÁS, *Ensayo sobre el Derecho a la propia imagen*, Madrid, Reus, 1931, p. 74.

¹³ THÉODORE IVAINER, “Le magnétophone, source ou preuve de rapports juridiques en droit privé”, *Gazette du Palais*, 1966, Paris, Doctrina, p. 95.

recorded and transmitted in multiple ways. Today the voice is already used to formalize contracts, to give orders to certain technological devices, or even to converse with them.

However, this is not without risk, as Zuboff points out. Today, voice is also a source from which to draw the “human experience” due to the increasing use of virtual voice assistants¹⁴. According to Zuboff, there is a strong case for continuing to rely on voice, as informal dialogues help to blur the boundaries between the assistant and the user. In the sociologist's words, “the more we feel like turning to the device as confidant, nanny, babysitter, governess and support system—a sort of incorporeal and omnipresent “Mrs. Doubtfire” for each person¹⁵—the more experience we will let it convert and transfer to it, and the richer its delivery operations will become. Communication is the first human pleasure, and a conversational interface has a special value because of the frictionless facility with which a simple emission of voice can provoke an action, especially if it is a mercantile action”¹⁶.

All this will force us to ask ourselves about the role of voice in contracting; about how consent should be formalized when interacting with applications that use voice recognition; or, finally, whether privacy policies are sufficiently guaranteed when dealing with our voices.

2. Voices and contracts: the voice in the different phases of contracting

Voice is an element that can be present in all phases of the contract, from negotiation to performance. However, its function will be completely different in some phases than in others. Thus, it can be a mere communicative vehicle between the parties, a means of formalization of the contract or constitute the object of the contract itself. Here we will start from the distinction traditionally adopted by the doctrine between the phases of generation, perfection and consummation of the contract to analyze the role that the voice can play in each of them.

¹⁴ Unlike the GDPR, the Directive does recognize the not inconsiderable value that personal data have acquired in the market economy, being collected in its art. 3. 1 that the Directive will be applicable not only to contracts in which the consumer pays a price, but also when he provides personal data to the entrepreneur, except when the personal data provided are processed exclusively for the purpose of providing the digital services, or to enable the entrepreneur to comply with the legal requirements to which he is subject and he does not process the data for any other purpose (an exception that reminds us of the provisions of art. 6.1. b) of the GDPR). Regarding the treatment of data as valid counter-performance in contracts for digital goods and services, see S. LOHSSE / R. SCHULZE / STAUDENMAYER (eds.), *Data as Counter-Performance - Contract Law 2.0?*, Baden-Baden, Nomos, 2020.

¹⁵ That this sociologist calls her a “lady” is not trivial, since there are already several studies that criticize the indisputable feminine identity of virtual voice assistants. We highlight here the one by MARK WEST / REBECCA KRAUT / HAN EI CHEW: “I'd blush if I could. Closing Gender Divides in Digital Skills Through Education”, 2019, *UN report*. Accessible at: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1> (29.08.2022). To try to fight against this stereotype of virtual assistant that perpetuates the image of women as assistants and caregivers, see the work of SOLEDAD TORRES GUIJARRO / MERCEDES RODRÍGUEZ GARCÍA/ IRIA VÁZQUEZ SILVA, “Desarrollando desde un enfoque igualitario: asistencia digital wikifeminista”, *Cátedra Feminismos 4.0*, Universidade de Vigo. Available at <https://catedrafeminismos.gal/wp-content/uploads/2020/09/poster-asistencia-Dixital-Wikifeminista-Catedra-Feminismos-40.pdf> (29.08.2022), in which they propose the development of a conversational agent with a gender perspective, with the idea of giving it a female scientist personality.

¹⁶ SHOSHANA ZUBOFF, *La era del capitalismo de la vigilancia*, Madrid, Paidós, 2020, pp. 351-352.

2.1. The voice in the initial phases of the contract

The initial phases of the contract comprise the preliminaries of the formation of the contract, which result in a series of acts from which the contractual consent arises¹⁷. These are acts that the interested parties carry out for the purpose of drawing up, discussing and concluding a contract, and may consist of both oral and written conversations or negotiations¹⁸.

In the case of the former, the voice will have a merely communicative function, as a vehicle of expression of the will to initiate the deals or conversations. What is relevant in this phase will be the content of what is expressed through the voice, especially with regard to compliance with the duties to negotiate in good faith, to provide certain information, the duty of protection and the duty of confidentiality¹⁹.

However, there is one issue that should be considered when we are dealing with oral conversations: the importance that the expression and intonation of the voice may have when expressing the will to contract, since it could be decisive when interpreting this will. Traditionally, two types of interpretations have been distinguished (which serve both for contracts already formed and to determine whether there is an intention to be bound), namely, the subjective interpretation, linked to the actual will of the declarant, and the objective interpretation, which seeks the meaning that can reasonably be found by the addressees of the statement²⁰. In both types, and although it will normally be a matter of interpreting the content of what was said²¹, it will also be necessary to take into account the intention that emerges from the use of the voice itself²²: if they have been manifestations made with irony (for example the *iocandi causa*²³), which can be deduced from a certain intonation; if they have been firm and taxative, declarations that demand a more defined and serious type of vocal expression; or, finally, if the expressive curves of the voice imply or not a will to contract, which will have to be analyzed according to each language and culture, since the expression

¹⁷ LUIS DíEZ-PICAZO Y PONCE DE LEÓN, *Fundamentos del Derecho civil patrimonial, I. Introducción. Teoría del contrato*, Cizur Menor (Navarra), Thomson-Civitas, 2007, p. 309.

¹⁸ LUIS DíEZ-PICAZO Y PONCE DE LEÓN, *Fundamentos del Derecho civil*, cit., p. 311, for whom a legal relationship is not created between the parties, but this does not mean that the conversations are not relevant, since they may have importance for the formation of the contractual will and for the interpretation of the contract.

¹⁹ Following the classification of LUIS DíEZ-PICAZO Y PONCE DE LEÓN, *Fundamentos del Derecho Civil*, cit. pp. 311-314. In turn, M^a PAZ GARCÍA RUBIO, *La responsabilidad precontractual en el Derecho español*, Madrid, Tecnos, 1991, pp. 43 et seq. divides the duties of the pre-contractual phase into duty of information, duty of loyalty (which includes the duty of reserve or secrecy) and duty of protection.

²⁰ LUIS DíEZ-PICAZO Y PONCE DE LEÓN, *Fundamentos del Derecho civil*, cit., p. 330.

²¹ For ÁNGEL CARRASCO PERERA, *Derecho de Contratos*, Cizur Menor (Navarra), Thomson Reuters-Aranzadi, 2010, and according to the first rule of contractual interpretation, art. 1281.I CC, it is presumed that the externalized statements express the real will or intention of the contracting parties.

²² We are aware that in addition to the phonetic or spoken language, when interpreting the intention of the message, it will be necessary to pay attention to the kinesic language (body movements that accompany the words), and even to the proxemic language (which takes into account the physical space in interpersonal relationships).

²³ The statement *Iocandi causa* ("because of a joke") is that made with lack of seriousness or in a humorous tone, without the intention of committing oneself. According to the second paragraph of art. 1281 CC, if the words seem contrary to the evident intention of the contracting parties, the latter will prevail over the former. ÁNGEL CARRASCO PERERA, *Derecho de contratos*, cit., p. 141, says that "if the addressee did not know the lack of seriousness of the declaration made by the other party, and the latter notices this fact, the declarant must immediately deny his own declaration, at the risk that the contract will be valid against him, as if his lack of negotiating intent were a simple mental reservation, irrelevant".

of the speech, together with the intention derived from the meaning of the words, is manifested differently even in regions that share the same language²⁴.

All the aforementioned is also useful when analyzing the function of the voice in the perfection of the so-called verbal contracts, one of the issues that we shall address below.

2.2. The voice in the phase of conclusion of the contract

At the contract conclusion phase, the contract can be concluded orally, where the voice will again be the protagonist, albeit with its communicative function; or it can also be used as a means of authenticating the signing of the contract, for which biometric voice recognition techniques can be used to identify the owner of the voice who is giving his contractual consent.

a) Verbal contracts

In a broad sense, the word “form” in relation to contracts has been used to designate anything that serves as a vehicle for the externalization of the contractual declaration of intent. In order to be legally relevant, the contractual will must be manifested by means of a “declaration” of will addressed to the other party²⁵. And this declaration could not be known if it is not expressed by means of some suitable sign, normally through social means of communication where the parties make use of the language through the oral or written word. Therefore, from this perspective, all contracts would be formal, since they all require some form to be entered into. In a narrower sense, form is everything that the law requires above and beyond the simple will of the party in order for a promise to be binding²⁶. Therefore, normally when we speak of the “form requirement” we are making the existence, effectiveness or proof of the contract dependent on the will having been externalized in a specific manner.

Generally, the doctrine has distinguished, based on this second meaning of the term “form”, formal contracts and non-formal contracts, the latter being those whose validity, perfection and effectiveness depend only on the existence of the consent of the contracting parties. Within the latter, it will not be unusual to find verbal contracts, which are those that have been agreed between the parties orally, without being formalized or recorded in any documentary support²⁷.

²⁴ A paradigmatic case is that of the Spanish language, in which the cultural aspects of each region/country affect both the accent and certain expressions or intonations that may vary the meaning of the same sentence.

²⁵ As explained by ÁNGEL CARRASCO PERERA, *Derecho de Contratos*, cit., p. 139, the mere contractual will of the parties, even if their purpose coincides, does not constitute contractual consent, since it needs to be transmitted by means of a “declaration of will” addressed to the other party.

²⁶ LUIS DíEZ-PICAZO Y PONCE DE LEÓN, *Fundamentos del Derecho civil*, cit., p. 287.

²⁷ Another issue will be to prove its existence, a matter that is usually solved by means of witnesses, acts, facts or documents that, although not being the contract itself, do prove that it exists.

The Spanish legal system does not place any obstacle to this type of contract, since art. 1278 CC —similarly to art. 219 Portuguese CC— allows “contracts to be binding regardless of the form in which they have been entered into, provided that they meet the essential conditions for their validity”. This precept contains the so-called “spiritualist principle”, traditional in our Law since the *Ordenamiento de Alcalá*, according to which validity is granted regardless of the form²⁸. The verbal form has been and continues to be the usual one in contracts of small economic amount, and even in some of greater entity.

b) The signing of contracts by means of the voice

Although Art. 1278 CC advocates freedom of form in contracting, the same CC establishes two exceptions: those contracts for which the law requires a certain form, usually written or notarized documentation as an unavoidable requirement for the contract to be concluded, or those in which the parties wish the contract not to be formalized until a certain document is signed. Both in the first case and in the second we can consider the possibility that the signature of such documents is done through the voice.

The signature of contracts was traditionally understood only as handwritten, but with technological advances, not only a digital signature is now allowed —which would be a digital substitute for the graphic and handwritten signature—, but also the signature by means of biometric recognition techniques, such as those based on the fingerprint, the iris of the eye, or the voice. Thus, there is already some discussion of the “voiceprint”, which will allow the voice “signature” to be integrated into an electronic document in the same way that a handwritten signature is incorporated into a paper document²⁹.

What is achieved through this signature, therefore, is an electronic identification of the person, in accordance with what Regulation (EU) No. 910/2014 understands as such in Art. 3: the process of using a person's identification data in electronic format that uniquely represents that person, and which in this case is based on his or her own voice. This technology relies on the measurement of phonetic and morphological parameters that make the voice unique to each person, just like the fingerprint. While it is true that the voice varies slightly over the years, these signature systems, through artificial intelligence, would learn these small variations each time the person uses his or her voice. It would be recorded by a microphone with integrated software that allows such analysis of voice parameters.

The companies that have implemented it so far activate the signature of the contract by voice once the parties have agreed to contract. This will require a platform that enables “biometric signing” by voice, where the contracting party records his or her voice reading the terms of

²⁸ LUIS DíEZ-PICAZO Y PONCE DE LEÓN, *Fundamentos del Derecho civil*, cit., pp. 290-291, for whom this consecration of freedom of form, an evident progress in the juridical order, favors human freedom in general.

²⁹ Regulation (Eu) No 910/2014 Of The European Parliament And Of The Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, establishes a legal framework for electronic signatures.

the agreement and personal data. The voice recording, along with the biometric parameters of it, are embedded in a PDF document, encrypted and sealed, and the signatory receives, at the same moment, a copy of the PDF document with his/her embedded voice signature, so that his/her voice and what he/she has signed with it can be heard, thus providing the same evidentiary capacity that the other contracting party has. Thus, each signature will be linked to a single document, guaranteeing the identification of the signatory³⁰.

This form of contracting can be done remotely, which has already begun to be used in the banking sector, in which the contracting party and holder of the voice signature can be recognized *a posteriori* thanks to the audio that the signature incorporates. This will allow contracts to be signed remotely without the need to be present in a bank office; however, consumer law requirements must be complied with when the signatory is a consumer, which will be the majority of cases. Consequently, it will be necessary that, both during the pre-signing and post-signing phases, the business person has complied with the obligations stipulated by articles 97 and 98 of the Spanish Consumer Act³¹ and by Law 34/2002³², on information society services and electronic commerce, and Law 22/2007, on distance marketing of financial services for consumers. Thus, before the consumer gives his consent by means of his voice, the bank must provide him with information regarding the essential characteristics of the goods and services, his identity, contact details, price, guarantees, right of withdrawal, and other requirements established in art. 97 of the Consumer Act. This information, which in this case will be provided in the telephone conversation, must also refer the consumer to another complementary means to expand the required information, such as the website of the entity, or the sending by email or regular mail of this information³³. After signing the contract, the bank must provide the consumer with confirmation of the contract concluded on a durable medium and within a reasonable period of time, including the pre-contractual information of art. 97 if it had not been previously provided on a durable medium. Otherwise, the contract could be annulled based on the action of art. 100 of the Consumer Act.

In short, contracts signed by voice must scrupulously comply with all the requirements of prior information and subsequent documentary confirmation, including the PDF document resulting from the voice signature process, not only the voice recording but also the terms and conditions of the contract, which are currently given to the consumer by the banks.

The problem we find in the new uses of voice as a signature of contracts is the small print of the privacy policies of these software developers that allow electronic signatures. At least we are aware that in one of them there is a subcontracting, for the provision of the signature service by voice, with one of the current technological giants, whose main activity is the sale of products of all kind. Therefore, as much as the privacy policy states that the subcontractor

³⁰ MARÍA BERMÚDEZ BALLESTEROS, "La protección del consumidor en la firma de contratos a través de la voz", *Publicaciones jurídicas Centro de Estudios de Consumo*, oct. 2019, p. 2.

³¹ The Spanish Consumer Act is contained in the *Real Decreto Legislativo 1/2007, de 16 de noviembre, por el que se aprueba el texto refundido de la Ley General para la Defensa de los Consumidores y Usuarios y otras leyes complementarias*.

³² *Ley 34/2002, de 11 de julio*.

³³ MARÍA BERMÚDEZ BALLESTEROS, "La protección...", cit., p. 4.

“shall also have the status of data processor” and “shall also be obliged to comply with the obligations established in this document for the Data Processor and the instructions issued by the Data Controller”, we believe that this information should be provided to the consumer in a more direct way, and not hidden as is the case in the infinity of terms and conditions of privacy³⁴.

In addition, these privacy policies do not specifically mention “voice data”, when it should be specified that it is the primary personal data to be processed. Voice is a biometric data, as it follows from Art. 4, paragraph 14 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR)³⁵. And the identification of a user through voice data involves the processing of biometric data, an issue that follows from the GDPR and that is mentioned in the Guidelines 02/2021 adopted on July 7, 2021 by the European Data Protection Committee on virtual voice assistants (section 3.4.3, paragraphs 81 and 82).

Therefore, in these contracts that have an embedded voice signature, the express consent of the owner of the voice for its processing should be required, especially when this processing also depends on a subcontractor company that, *a priori*, has nothing to do with the object of the contract being signed. This is because art. 9.1 GDPR prohibits the processing of, among others, “biometric data intended to uniquely identify a natural person”, unless express consent is given (art. 9.2. a) GDPR).

3. User interaction by means of voice and virtual voice assistants

In this section we will discuss the clauses inserted in digital content supply contracts that refer, in a generic way, to “voice data” or “interaction with the service through voice commands”, present in platforms such as Netflix, in social networks such as Facebook, or in operating systems such as Microsoft. We will also analyze those inserted in the virtual voice assistants themselves, such as the well-known Siri (Apple's Siri) or Google Assistant³⁶.

³⁴ In this regard, see FirVox Mobile's Privacy Policy, which establishes Amazon Web Services as a subcontractor, “which will also have the status of processor” of personal data, in this case voice data. Available here: <https://biometricvox.com/politica-de-privacidad-firvox-mobile/> (09.09.2022).

³⁵ Art. 4.14 Regulation (EU) 2016/679 defines “biometric data” as personal data obtained from specific technical processing, relating to the physical, physiological or behavioral characteristics of a natural person which allow or confirm the unique identification of that person, such as facial images or fingerprint data. The definition given by the GDPR is rather narrow, as it only mentions two characteristics - facial images and fingerprints - by way of example. Much more illustrative is the delimitation reflected in Opinion 4/2007 of the Art. 29 Working Group, on what biometric data are: biological properties, physiological characteristics, personality traits or tics, which are at the same time attributable to a single person and measurable, even if the models used in practice to measure them technically imply a certain degree of probability. Typical examples of biometric data are provided by fingerprints, retinal patterns, facial structure, voices, but also hand geometry, venous structures and even a certain deep-seated skill or other behavioral characteristic (such as handwriting, a particular way of walking or talking, etc.).

³⁶ See, for example, the privacy statement of Google Assistant: <https://policies.google.com/privacy?hl=es> (09.09.2022).

3.1. Voice processing as biometric data

First of all, it could be argued that the function of voice in these services is that of a communicative instrument. Therefore, it would be limited to being a form of expression of the will without there being a treatment of the voice as personal data. Thus, voice would be the vehicle through which we would execute a function in the application, a function that could also be implemented through, for example, the manual selection of a command. However, we are of the opinion that since the voice is recorded in these devices (as could happen with the image in the case of facial recognition systems), and since the providers of these services have the means to identify the person whose voice is captured, the existence of voice processing is indisputable. This is also in line with the Guidelines 02/2021 on virtual voice assistants.

Secondly, the very classification as “voice data” by the providers of these services implies an implicit admission of its processing. In “privacy policies” or “consent to the processing of personal data” it is not uncommon to find clauses along the following lines: “For example, we will manually review short excerpts of voice data where we have followed steps to uninstall and enhance voice recognition technologies. It is possible that employees of [the company] or vendors who are working on behalf of [the company] may perform this manual review”³⁷.

3.2. Analysis of the consent given by the users to the processing of their voice data

a) When the processing of voice data is necessary for the performance of the contract. The case of virtual voice assistants

As a premise, we must point out that art. 3.1 of Directive 2019/770 exempts from its scope of application contracts in which, although personal data must be provided in exchange for the service, they are processed exclusively for the purpose of providing the digital services, or to enable the trader to comply with the legal requirements to which he is subject, not being allowed to process the data for any other purpose.

It is recital 25 of the Directive that deals, in part, with these exceptions, although further exemplification of situations would be welcome. It lists three examples of non-application: (i) cases where registration of the consumer is required under the applicable law for security and identification purposes; (ii) situations where “the trader collects only metadata such as information about the consumer's device or browsing history, except where this situation is

³⁷ Privacy statement of Microsoft, under the section “How do we use your personal data”. Available here: <https://privacy.microsoft.com/es-es/privacystatement> (09.09.2022).

considered a contract under national law” and (iii) situations where “the consumer, without having concluded a contract with the trader, is exposed to advertisements exclusively in order to gain access to digital content or services”. We believe that it would not have been superfluous to add, as the doctrine has already said, some cases in which data are collected exclusively to provide digital content or services beyond the metadata mentioned, as would be the case of navigation services, which need to know location data for the sole purpose of recommending the route to the consumer³⁸.

Following this line, it could be argued, therefore, that the clauses relating to the processing of “voice data” could be essential for the operation of the voice assistance feature of audiovisual service platforms, or of computer operating systems. Thus, the processing would be lawful under Art. 6.1. b) GDPR, according to which “the processing is necessary for the performance of a contract to which the data subject is a party”, as the user would have given consent to the use of that service, in which the use of voice would be included. In fact, the concept of “voice interaction” could serve as a basis for this argument, understanding that we are dealing more with the use of voice as a tool for communication, for interaction, for expression of will. In other words, what is really important would be the content of what is said by voice, the order given by the user to the machine, and which would be used as a mere vehicle of expression.

However, we do not believe that this position is correct. Voice assistants, especially those used by personal devices such as cell phones, often ask the user to pronounce a series of phrases for their configuration, so that the assistant recognizes that particular voice and not another (the so-called “voice recognition”, with which voice models or “voiceprints” are created). In other words, the treatment of the voice, as a form of identification of the person, is clear in these cases. This leads us to argue that the interpretation of Art. 6.1 b) should be more restrictive when we are dealing with biometric data, as would be the case with voice. One of the innovations of the GDPR is precisely that biometric data are added to the special categories of data. Thus, art. 9.1 GDPR prohibits the processing of, among others, “biometric data intended to uniquely identify a natural person”, unless express consent is given (art. 9.2. a) GDPR). Therefore, the voice, indisputably biometric data, will always require, for its processing, the express consent of the holder, also for the case where such processing of the voice is an “accessory activity”, but necessary, for the execution of the contract. In other words, “voice data” clauses would have to be expressly and individually accepted by the user. Again, we understand that this statement is in line with the aforementioned “Guidelines 02/2021 on virtual voice assistants”.

If we look at the privacy policies of virtual assistants, we do not believe that they currently contain clear clauses explaining the specific treatment of personal voice data, nor do they

³⁸ Example mentioned by A. METZGER: “A Market Model for Personal Data: State of Play under the New Directive on Digital Content and Digital Services”, *Data as Counter-Performance – Contract Law 2.0?*, Baden-Baden, Nomos, 2020, pp. 25-45, esp. p. 37.

specify their consideration as specially protected data, as is the case with all biometric data³⁹. According to the Guidelines 02/2021 on virtual voice assistants, the assistants should comply with a series of requirements —for the time being only recommendations—, which can be summarized as follows: (i) the “voiceprints” generated should be stored exclusively on the local device, and not on remote servers; (ii) both standardized rules (e.g. ISO/IEC 24745) and biometric model protection techniques should be applied; (iii) identification should be guaranteed to be sufficiently accurate to reliably associate personal data with the correct subjects; and (iv) identification accuracy should also be guaranteed to be similar for all user groups, ensuring that there are no substantial biases towards different demographic groups.

Analyzing two of the most widely used assistants (Siri, from Apple, and Google Assistant, from the homonymous company), we can see that in the last two years or so, both have taken care to implement some of these recommendations, but in our opinion not sufficiently. For example, we see that in the case of Siri it is said that the priority is to process the data on the device itself⁴⁰, but in the case of Google Assistant it is said that “Google stores the data from your interactions with Google Assistant on its servers, which are in its data center”⁴¹. Both allow the user to delete conversations and requests made to the assistant, or to limit the time they are kept.

However, neither of the two shows that they are dealing with biometric data, such as voice, and therefore with special data that need greater protection. In short, we see that they focus more on the protection of the content of what is said than on the medium itself, the voice, which, as biometric data, would also need such protection. This would be guaranteed, among other ways, by the application of those standardized norms for the protection of biometric models (ISO/IEC 24745, among others) recommended by the Guidelines, but which, of course, are not mentioned in any of the privacy policies studied.

b) When the processing of voice data is not necessary for the performance of the contract. The case of “voice data” clauses in other services

In this section we will refer only to voice data clauses present in services whose main function is not to be virtual voice assistants (as could be, for example, a Smart TV).

³⁹ We will leave for another time the criticism that we believe that all these privacy policies deserve. Not only because of their lengthy and confusing wording, already criticized by others for years, but also because of their current tendency to hide them among interfaces and links that, under the request “if you want to know more about... click here”, turn the search for any rule into a nightmare. In fact, we believe that it used to be easier to find certain policies before (even if you had to read a hundred pages) than now, where companies use these “treasure hunt games” that only confuse the user even more.

⁴⁰ On this matter, see “Privacy and Siri evaluation program”: <https://support.apple.com/es-es/HT210558> (09.09.2022).

⁴¹ In this regard, see, “Security and data privacy in devices that work with the Google Assistant”: <https://support.google.com/googlenest/answer/7072285?hl=es#zippy=%2Cd%C3%B3nde-se-guardan-mis-datos> (09.09.2022).

If it were argued that the clauses relating to “voice data” are not necessary for the performance of the contract, to analyze the lawfulness of the processing of this data we will have to turn to art. 6.1. a) GDPR. Thus, we will have to ask ourselves whether the data subject (consumer for the purposes of Directive 2019/770) has consented to the processing of this data “for one or more specific purposes”, and always taking into account the parameters set by arts. 4.11 and 7 GDPR. It follows from them that consent refers to “any freely given, specific, informed and unambiguous expression of will by which [the data subject] agrees, either by a statement or by a clear affirmative action, to the processing of personal data concerning him or her”.

The difficulty will therefore lie in the interpretation of these indeterminate legal concepts, since the question that arises is whether we have free will when, in order to contract the service, we must accept the processing of personal data—in our case, the voice— whose usefulness/purpose does not seem essential for the operation of the contracted service, for example, a Smart TV. Does the fact that the uses to which our data will be put in lengthy and rambling texts mean that the consent has been given in an informed manner? If our answers were negative, we would be considering the consent as invalid, and at the same time delegitimizing a multitude of practices of companies providing these services, so it will be necessary to find a mid-point.

In the Italian doctrine, Resta has proposed, in accordance with the legislator and the interpretation that the Italian courts have made of consent for the last two decades, that we will not be dealing with free consent when the manifestation of will is formed under conditions of psychological pressure linked to situations of vulnerability or structural asymmetries of power. It even recalls that the courts have prohibited situations in which the provision of an online service is conditional on consent to the processing of personal data for purposes additional to the true object of the contract⁴².

In fact, this position on consent links with what is now said in Art. 7(4) GDPR which stipulates that “when assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract”. In order to combine this article of the Regulation (which, strictly interpreted, would end with all the reality that legitimizes the personal data market) with the provision of the Directive on the supply of personal data in exchange for a service, it has been argued that the rule of the former does not introduce a prohibition, but a mere parameter of assessment; or it could even be a presumption of invalidity of the consent given, but it is clear that it is a presumption *iuris tantum* and not *iuris et de iure*⁴³. However, the problem lies in the fact that the current practices of service providers, if they are similar in any way, it is in their opacity; so Resta stresses the need for commercial practices to be more transparent and specific as far as consent is concerned, concluding that what would be desirable would be a

⁴² GIORGIO RESTA, “I dati personali oggetto del contratto. Riflessioni sul coordinamento tra la Direttiva (UE) 2019/770 e il Regolamento (UE) 2016/679”, *Annuario del Contratto*, 2019, pp. 135-136.

⁴³ GIORGIO RESTA, “I dati personali”, cit., p. 140.

“granularity of consent” given, allowing the user to access the service without authorizing the subsequent processing of their data. In addition, the logical order would be first the provision of consent to the processing of personal data, and then the manifestation of the contractual will: differentiated but united acts, so that in the event of a revocation of consent to the processing of data (ex. art. 7.3 GDPR), there would be a termination of the contract⁴⁴.

In short, we believe that current practice does not allow us to affirm that consent has been validly given. If we take Microsoft's privacy policy as an example, we see that the statement on voice data collection is diluted among the rest of clauses that provide contractual consent to access the service⁴⁵.

4. Final conclusion

Voice is a personality feature that identifies and individualizes us. Moreover, as biometric data, it can be processed and recognized using biometric techniques. This has implications when using the “voiceprint” as a form of electronic signature for certain contracts; or when consenting to the processing of our voice data in services that use it, or in virtual voice assistants. In all these cases, we will have to be very careful when it comes to formalizing consent for the processing of voice data.

Currently, we believe that privacy policies and ways of expressing consent for such uses of voice do not sufficiently guarantee the rights of voice owners. In the case of voice signature contract signing, we detect loopholes in informing the user about potential subcontractors of electronic signature mechanisms, which appear hidden in the terms and conditions of the voice signature developer. In the case of voice virtual assistants, we believe that the standardized rules for the protection of biometric models recommended by the European Data Protection Agency have not yet been implemented by the companies developing these assistants. And in the case of clauses that provide for user interaction by voice on certain devices such as Smart TVs, we believe that there should be a “granularity of consent” to be able to prevent, where appropriate, the processing of voice data when that specific service is not desired.

Bibliography

AMMERMAN YEBRA, JULIA, *El derecho a la propia voz como derecho de la personalidad*, A Coruña, Colex, 2021

⁴⁴ GIORGIO RESTA, “I dati personali”, cit., pp. 141-143.

⁴⁵ Available at <https://privacy.microsoft.com/es-es/privacystatement> (09.09.2022).

- BERMÚDEZ BALLESTEROS, MARÍA, "La protección del consumidor en la firma de contratos a través de la voz", *Publicaciones jurídicas Centro de Estudios de Consumo*, oct. 2019
- CARRASCO PERERA, ÁNGEL, *Derecho de Contratos*, Cizur Menor (Navarra), Thomson Reuters-Aranzadi, 2010
- CORNUT, GUY, *La Voix*, Paris, Presses Universitaires de France, 1990
- DE CUPIS, ADRIANO, *I diritti della personalità*, Vol. IV, T. I, Milán, Giuffré, 1959
- DÍEZ-PICAZO Y PONCE DE LEÓN, LUIS, *Fundamentos del Derecho civil patrimonial, I. Introducción. Teoría del contrato*, Cizur Menor (Navarra), Thomson-Civitas, 2007
- FERRÁNDIZ AVENDAÑO, PABLO, "Protección del artista frente a la imitación de su voz", *Actualidad Civil Jurisprudencia*, no. 21, Wolters Kluwer, 2010
- GARCÍA RUBIO, M^a PAZ, *La responsabilidad precontractual en el Derecho español*, Madrid, Tecnos, 1991
- GUIMARÃES, RAQUEL, "A tutela da pessoa e da sua personalidade: algumas questões relativas aos direitos à imagem, à reserva da vida privada e à reserva da pessoa íntima ou direito ao carácter", *A tutela geral e especial da personalidade humana* [ebook], Centro de Estudos Judiciários, 2017
- HUET-WEILLER, DANIELLE, "La protection juridique de la voix humaine", *Revue Trimestrielle de Droit civil*, 1982
- IVAINER, THÉODORE, "Le magnétophone, source ou preuve de rapports juridiques en droit privé", *Gazette du Palais*, París, Doctrina, 1966
- LOHSSE, S. / SCHULZE, R. / STAUDENMAYER (eds.), *Data as Counter-Performance - Contract Law 2.0?*, Baden-Baden, Nomos, 2020
- METZGER, A., "A Market Model for Personal Data: State of Play under the New Directive on Digital Content and Digital Services", *Data as Counter-Performance - Contract Law 2.0?*, Baden-Baden, Nomos, 2020
- RESTA, GIORGIO, *Autonomia privata e diritti della personalità*, Nápoles, Jovene Editore, 2005
- RESTA, GIORGIO, "I dati personali oggetto del contratto. Riflessioni sul coordinamento tra la Direttiva (UE) 2019/770 e il Regolamento (UE) 2016/679", *Annuario del Contratto*, 2019
- RUIZ Y TOMÁS, PEDRO, *Ensayo sobre el Derecho a la propia imagen*, Madrid, Reus, 1931
- TORRES GUIJARRO, SOLEDAD / RODRÍGUEZ GARCÍA, MERCEDES / VÁZQUEZ SILVA, IRIA, "Desarrollando desde un enfoque igualitario: asistencia digital wikifeminista", *Cátedra Feminismos 4.0*, Universidade de Vigo. Available at <https://catedrafeminismos.gal/wp-content/uploads/2020/09/poster-asistencia-Dixital-Wikifeminista-Catedra-Feminismos-40.pdf> (29.08.2022)

SCHIERHOLZ, ANKE, *Der Schutz der menschlichen Stimme gegen Übernahme und Nachahmung*, Baden-Baden, Nomos, 1998

SCHOLES, PERCY A. / OWEN WARD, JOHN / DEVOTO, Oxford Dictionary of Music, T. I and II, Barcelona, Edhasa / Hermes / Sudamericana, 1984

WEST, MARK / KRAUT, REBECCA / CHEW, HAN EI, "I'd blush if I could. Closing Gender Divides in Digital Skills Through Education", 2019, *UN report*. Accessible at: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1> (29.08.2022)

ZUBOFF, SHOSHANA, *La era del capitalismo de la vigilancia*, Madrid, Paidós, 2020

(texto submetido a 22.09.2022 e aceite para publicação a 27.09.2022)